

基于电商大数据的农产品短期经营风险预测研究

——以家庭经营梨果种植户为样本

程欣炜,岳中刚

(南京邮电大学经济学院,江苏南京210023)

摘要:农产品电商具有农业经营改善和农户福利提升的积极作用,在电商参与模式下识别和预测农产品短期经营风险有助于农户家庭主动管理和积极干预电商嵌入结构,也可为区域电商布局优化提供理论依据和预警方案。文章以江苏省3755户家庭梨果种植户为样本,采集种植投入、电商交易和借贷保险等电子台账大数据,对比分析线性计量模型、一维非线性机器学习模型和二维深度学习模型对农产品短期经营风险结果和风险决策的预测能力。结果表明,电子台账大数据包含预测农产品短期经营风险的信息含量,基于特征灰度的二维卷积神经网络具有精准而稳定的预测能力,电商交易的预测信息含量高于种植投入和借贷保险且三者相互印证。在此基础上,从电商数据规范、台账软件升级、电商咨询建设和数字平台管理角度提出建议。

关键词:农产品经营;风险预测;电商大数据;卷积神经网络;计算机视觉

中图分类号:F713 **文献标志码:**A **文章编号:**1000-2154(2022)09-0016-14

DOI:10.14134/j.cnki.cn33-1336/f.2022.09.002

Short-term Agricultural Operating Risk Prediction with Big Data of E-commerce: A Case Study of Family-run Pear Farmers

CHENG Xinwei, YUE Zhonggang

(School of Economics, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

Abstract: Agricultural e-commerce has a positive effect on operation improvement and welfare enhancement. Identifying and predicting its risks helps to actively manage and intervene embedding patterns, and also provides theoretical basis and early warning solutions for regional e-commerce optimization. Electronic ledger data of production, transaction and finance of 3755 pear farmers in Jiangsu Province were collected to compare the predictive capability of linear model, 1-dimensional non-linear machine learning models and 2-dimensional deep learning model on the risk outcomes and decisions of agricultural operations. The results show that ledger data have adequate information content to predict short-term agricultural risk. Convolution Neural Networks on feature grey scale has the best prediction ability. And the prediction information of transaction data is higher than production and financial data with mutually validated structure. Recommendations are made from the perspectives of e-commercial data specification, ledger software upgrading, consulting service construction and digital platform management.

Key words: agricultural operation; risk prediction; big data of e-commerce; convolution neural networks; computer vision

收稿日期:2022-04-14

基金项目:教育部人文社会科学基金一般项目“基于深度学习的农产品电商风险预警和金融对策研究”(18YJC790016);江苏省社会科学应用研究精品工程课题“脱贫攻坚阶段农产品电商对小农户有机衔接现代农业发展效率的影响研究”(20SYC-057)

作者简介:程欣炜,男,讲师,经济学博士,主要从事农业经济和农业人口统计研究;岳中刚,男,教授,管理学博士,主要从事数字经济研究。

一、引言

数字化、智能化、精准化是未来农业生产经营的重要目标。《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》强调加快发展智慧农业,推进农业生产经营和管理服务数字化改造;2022年中央一号文件则明确提出推进数字乡村建设和智慧农业发展,以数字技术赋能乡村公共服务,拓展农业农村大数据应用场景。然而,《中国农业风险管理发展报告(2021)》指出,近年来我国除自然风险和市场风险外的非传统农业风险因素日益增多,农业生产经营应对突发事件和短期风险的监测预警系统和手段集成较为滞后,导致农业风险防范的科技支撑力量有限、科技信息应用不足。一方面,学界认识到农产品短期经营风险的识别和预测对事前风险管理的积极意义,特别是对低收入农户的主动风控意愿(陈新建和韦圆圆,2019)^[1]和新型职业农民的持续务农意愿(沈琼,2020)^[2]具有关键性影响;而另一方面,尚燕和熊涛(2020)^[3]采用东北三省和湖北省数据证实了异质性农业风险的管理错配,农户对市场经营风险的管理意愿远远高于自然风险,但当前的风险识别和管理技术主要针对自然风险,农户对市场经营风险的管理效率较低,亟须引入大数据来提升事故识别和损失程度的预测精度(丁少群等,2021)^[4]。正如Mendras较早年在*La Fin des Paysans*中指出农户对采纳现代技术的审慎态度,国内数据证实禀赋依赖下的农业行为存在风险中性(王志刚等,2005)^[5],即农户缺乏利用下一个种植周期的经营预期来调整生产经营行为的技术和能力;而当冬歇过后盘点收入损失或比较成本超出心理阈值时,农户将转向流出土地和非农就业,从而造成下一个种植年的聚集性撂荒现象(谢花林和黄莹乾,2022)^[6]。因此考虑到农业经营具有明确而稳定的周期性特征,农户需要培养以种植年为单位的短期风险结果认知能力(包括营收、清库和偿债等三个方面),并以风险结果为基础建立动态风险决策机制(包括电商选择和持续务农等两个方面)。

随着通信技术发展和互联网普及,农产品电商发展在农业经营改善和农户福利提升方面的积极作用被广泛证实。农产品电商降低或消除了农产品交易成本(Porter,2001)^[7],提升了家庭农业收入和现金收入(Khanal等,2016)^[8],借助市场驱动电商资本下沉对农户增权赋能(周浪,2020)^[9]。早在“十五”计划中,中央就要求重视信息网络技术在农产品交易、农业技术推广等领域的积极应用,2014年后农村电商成为我国精准扶贫十大工程之一,在脱贫攻坚阶段获得有效推广;《数字乡村发展行动计划(2022—2025)》要求深化农产品电商发展,并在专栏中强调了深入推进“互联网+”农产品出村进城工程的重要性。但作为新兴的产销模式,电商组织的供应链更易受到市场冲击(Dutta等,2019)^[10],新的风险因素和防范需求被引入农业生产经营领域。从风险主体角度看,农产品电商打破了传统农业的议价体系并建立起渠道风险更高的销售模式,参与主体间博弈须借助监管加以控制(Zhang和Wei,2022)^[11],因此其数据赋能具有基于数字成本的竞争性,成本分摊结构受竞争主体数据利用效率的影响,均倾向于分摊数据利用效率较高的参与方的成本(肖迪等,2021)^[12]。而从风险感知角度看,农户对电商风险的感知会明显抑制线上销售意愿(汪兴东和刘雨虹,2021)^[13]和电商创业意愿(王艳玲和张广胜,2021)^[14],对数字技能较强的新生代农户影响显著,生产资料市场风险和网络基础设施建设则具有外生性(向丽和胡珑璞,2019)^[15];同时,强弱不均等的电商合作受感知关系风险的显著影响,而对感知绩效风险并不敏感(郝鸿等,2015)^[16],因此在营商风险和规则驯化下,小农户更易退出电商市场并导致农村资源流失(聂召英和王伊欢,2021)^[17]。

电商环境下农产品与工业品的经营风险存在类型结构和感知需求上的本质差异。以种植户为例,农产品电商的短期经营风险来自一个生产周期内种植、加工、销售和回款等多个经营环节的风险叠加,农户可以通过在全部或部分农业经营环节上退出电商参与而不影响其农业经营链条的完整性,也可以通过全部或部分退出农业生产而不影响其电商参与活动的完整性,因此农户的异质性风险感知成为维持农产品电商生态多样性的关键因素,导致传统的农业经营风险预测方法失灵。但嵌入电商环境也为建立新模式下的风险预警提供了坚实的数据和技术基础:一是电商形成并记录海量数据,具有反映和预测经营绩效的信息含量(廖理等,2021)^[18]。Kamble等(2019)^[19]对84项研究的元分析证实,数据驱动的电商农业供应链能够有效解决产业化不足、管理低效、信息模糊等问题;Dutta等(2019)^[10]则指出对电商数据的持续跟踪能够有效

控制风险。二是对电商大数据的非线性处理具备较为完善的技术条件,机器学习被广泛运用于农业经营和
 风险预测领域,特别是支持向量机和神经网络的应用(林明等,2013^[20];王骞,2011^[21]),部分学者将电商
 大数据引入农业分析,尝试论证自然语言处理(谢星雨和余本功,2021^[22];岑咏华等,2021^[23])和图神经网络
 (胡春华等,2021)^[24]对复杂电商数据的处理效果。

本研究基于对江苏省3755名稳定参与电商的家庭梨果种植户的调查结果,对比不同模型下电商大数
 据对农产品短期经营风险的预测能力,证实电子台账数据具有识别下一年度农产品经营风险结果和风险
 决策的信息含量,并通过建立基于特征灰度的二维卷积神经网络揭示风险预测的信息侧重。研究的边际贡
 献在两个方面:第一,从数据样本角度看,研究采集农户电子台账数据并将行为信息结构化,证实高频电子
 台账数据较低频经营特征数据具备更高的风险预测信息含量,突破以销售反馈(王颖和阮梦黎,2018)^[25]
 或金融借贷(徐鲲等,2021)^[26]为数据基础的简单评估体系;第二,从研究方法角度看,研究发展了 Dai 等
 (2018)^[27]认为传统预测模型并不适用于电商数据预测的结论,深入对比了线性预测、一维非线性机器学
 习和二维深度学习的风险预测能力,借鉴计算机视觉的可解释性理论,证实种植投入、电商交易和借贷保
 险等经营行为具有不同预测信息含量且相互印证。

二、理论分析与研究设计

(一) 理论分析

考虑到农业在抗自然风险和抗社会风险的能力上具有天然的弱质性,特别是小农户无法形成抵御市
 场风险的强大力量(刘艳,1998)^[28]，“安全第一”成为农户经济行为的拇指规则(Roumasset,1976)^[29]，对
 农业经营风险的准确识别和预测将有效引导农户灵活调整产销行为。如图1所示,研究以“信息含量—粒
 度提升—模型选择—信息贡献”为思路建立农产品短期经营风险预测理论框架,即在划分和明确农产品
 短期经营风险类型的基础上,首先对比农户经营数据和电子台账数据在风险预测上的有效信息含量,其次
 借助农产品电商的行为数据验证机制将特征变量从年粒度细化至周粒度,接着对比三类模型的农产品短
 期经营风险预测精确性和稳健性,最后细分电子台账数据的信息维度并证实其信息贡献和印证关系,从而
 科学构建基于电商大数据的农产品短期经营风险预测系统。

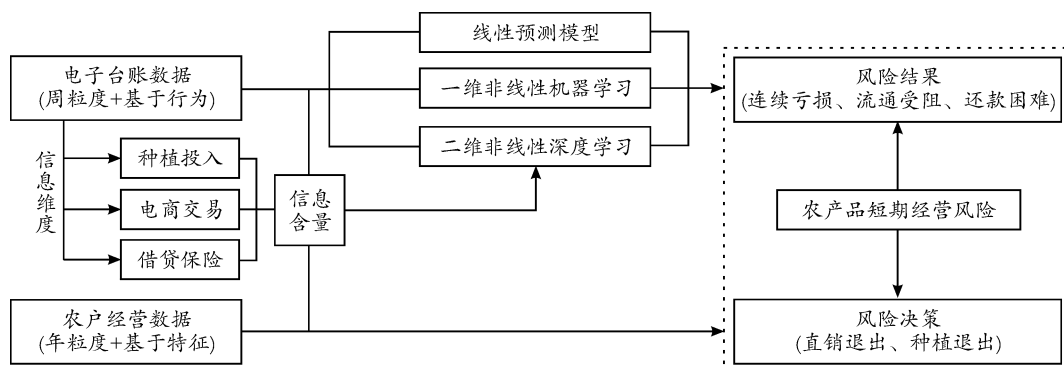


图1 研究的理论框架

在电商环境下,农产品短期经营风险可划分为风险结果和风险决策两个方面。就风险结果而言,习近
 平总书记在2016年4月的农村改革座谈会上强调“农民小康不小康,关键看收入”,当农业经营的收入预期
 不明确时,赊购农资农具的资金风险上升,农户将避免或缩减长期投入以便对下一年度的营商环境积极调
 整,因此连续亏损是评价农产品短期经营风险的关键指标;对于林果类农产品种植户而言,流通受阻会带
 来清货困难和果品降级。农业农村部《农产品产地流通及“最先一公里”建设调研报告》指出,我国果蔬和
 薯类农产品的产后损失率高达15%~25%,远低于发达国家的5%,造成每年近2亿吨的农产品浪费,特别

是应季农产品的流通饱和给种植户带来巨大的经营风险;而已有研究较少在收入资金回笼滞后、销售订单兑现困难等受压市场条件下探讨农户融资可得性,农户在上一个生产经营周期积累的还款压力往往聚集在种植年年末,缺乏流动性和有效抵质押品的普通农户很难融通到足够资金来维护自身在农村社会网络中的必要信用。就风险决策而言,传统农业经营的风险积累过程具有较强的隐蔽性,经营惯性使农户较少在第一产业内部调整就业结构,因此已有研究主要关注兼业劳动和土地流转决策,但电商环境下农户有机会在不脱离第一产业的条件下进行就业细分,如选择脱离种植而进入通勤的农业服务行业,或脱离直销成为标准化电商供应链的生产加工部门。

然而对农产品经营风险的预测并非易事,主要存在三个难点:第一,传统农业调研成本高,往往以年为时间单位构建基于特征观测的农户经营数据,造成当年数据仅能评价当年风险结果而无法预测未来风险,学界也尚未给出能够用于预测农业经营风险的指标或模型。但农产品电商形成并记录的大量行为数据具有交叉验证性,使农产品经营风险的跨年短期预测成为可能。第二,由于农业经营周期长,多季产品经营周期重叠复杂,金融行为能力与农户的非生产性禀赋相关,因此风险预测难度远大于工业企业,传统计量模型无法对非线性结构的风险模式进行分类。而农产品电商数据粒度更高,能够将时间维度和测度维度上的关联性状表现出来。第三,文献揭示了农户的风险感知对经营行为的影响机制,因此将农产品经营风险划分为风险结果和风险决策后,对风险决策的预测信息含量和模型适用的要求将明显高于风险结果。综合以上分析,研究提出假说 H1a 和 H1b:

H1a: 基于行为的电子台账数据较基于特征的农户经营数据更能准确预测农产品短期经营风险。

H1b: 电子台账数据包含农户经营数据中的短期风险预测信息含量。

在明确农产品短期经营风险预测的特征信息需求基础上,采用线性计量模型处理电商大数据将表现出两方面劣势:一是高频采样数据存在偶发参数问题,预测模型失去测度上的扩展性;二是遗漏非线性风险模式,在离散分类的预测上强调无偏性和可解释性,从而降低了预测精度。因此,非线性机器学习(ML)更适合预测农产品短期经营风险,包括风险结果和风险决策;其中,风险决策建立在不同农户对风险结果的模糊感知和反馈基础上,单一过程变量的调整并不能直接指向决策,必须通过农户在其他行为上所表现出的一致性风险态度加以衡量。机器学习模型能够有效识别电商大数据中由多个变量所形成的行为模式,并在异质性风险感知下进行预测,从而提升预测的精准性和稳定性。但机器学习算法将多种测度压缩为一个长向量进入模型,向量内元素独立,当风险模式同时存在时间和测度两个维度时,机器学习无法识别模式的空间不变性。将台账信息整理为“时间—测度”二维数据并按行执行归一化,可构建基于特征灰度的卷积神经网络(FG-CNN),借助计算机视觉理论模拟风险模式在水平方向的平移不变性和在水平、非水平方向的拉伸不变性,从而提升预测精度。同时,卷积神经网络是具有较多隐藏层的深度学习模型,引入池化层后参数量减少、感知野增加,将提升复杂风险决策(如种植退出)的预测稳定性。基于上述分析,研究提出假说 H2a 和 H2b:

H2a: 非线性机器学习对农产品短期经营风险的精预测度高于线性计量方法。

H2b: 基于计算机视觉的二维深度学习对农产品短期经营风险的预测准确性或稳定性高于机器学习。

引入农产品电商模式后,农户以更低的经济和社会成本参与区域内农产品供应链各个环节,特别是对弱势农户具有劳动激励作用(程欣炜和岳中刚,2021)^[30]。已有研究主要将传统农业经营模式下的风险识别集中在种植环节,强调自然风险下的成本损失和产量损失,在保险实践中表现为单位重量下的价格保险保费远低于收入保险,特定地区以少量鲜果品种为标的的价格指数保险试点并未获得全面推广普及,鲜果种植实践则仍以果树产量和死亡保险为主,且对投保农户的种植规模有要求,小农户存在参保门槛。而事实上,农户经营收入风险主要在交易环节兑付,同时这一环节上的经营损失无法获得种植补贴和保险赔付,因此研究认为电商环境下的农业经营数据存在种植投入、电商交易和借贷保险三个资金维度,农产品电商通过扩展农户销售渠道提升农户应对既有环境风险下的控制能力,体现了农户家庭突破当前资源禀赋和技术约束的行为意识和经营水平,较种植投入等生产信息和借贷保险等金融信息更具根本性和预期性。同时,考虑到电商大数据是家庭农业经营能力在多个经营行为层面的一致性体现,在测度上相互印证

并形成特定行为模式,因而同时引入种植投入、电商交易和借贷保险三类数据的预测能力将优于采用任意一类或两类数据。由此,研究提出假说 H3:

H3: 电子台账信息中的电商交易数据具有高于种植投入和借贷保险的风险预测信息含量,且三类数据能够相互印证而提升整体预测能力。

(二) 研究设计

基于上述理论分析框架,研究以年为时间单位采集农户经营数据(FMD),主要包含生产特征(ST)、电商特征(DT)和环境特征(HT);以周为时间单位采集电子台账数据(ESB),主要包含种植投入(PD)、电商交易(SD)和借贷保险(FD)。采用0~1二项变量表示农产品短期经营风险(AER)并划分为风险结果和风险决策,其中,风险结果包括连续亏损(LK)、流通受阻(LS)和还款困难(HK),风险决策包括直销退出(ZT)和种植退出(ST)。

1. 农户经营数据和电子台账数据的信息含量对比。以农产品短期经营风险(AER)为因变量,逐步引入农户经营数据(FMD)和电子台账数据(ESB)构建二项 $logistic$ 模型,如式(1)和式(2)所示。

$$AER_k = Sigmoid(\beta_0 + \beta_{1i} FMD_i + \varepsilon) > 0.5 \quad (1)$$

$$AER_k = Sigmoid(\beta_0 + \beta_{1i} FMD_i + \beta_{2j} ESB_j^y + \varepsilon) > 0.5 \quad (2)$$

其中, AER_k 表示5项经营风险指标($k=1,2,\dots,5$), FMD_i 表示9项农户经营指标($i=1,2,\dots,9$), ESB_j^y 表示15项电商台账指标 ESB_j 的年汇总值($j=1,2,\dots,15$)。引入 ESB_j^y 后,在式(1)中显著的 FMD_i 估计系数在式(2)中不再显著,则表明电子台账数据(ESB)包含或部分包含了农户经营数据(FMD)在风险预测上的信息含量。在回归模型基础上,研究分别构建基于农户经营数据、电子台账数据和两类数据的 $logistic$ 预测模型,从准确性角度验证其预测能力,并作为后续预测模型的基准结果。

2. 基于电子台账数据的预测方法对比。在确定电子台账数据具有风险预测信息含量的基础上,研究将数据集按2:1:1划分为训练集、验证集和测试集,分别构建一维机器学习模型和二维深度学习模型进行风险预测能力的对比分析。其中,一维机器学习选用支持向量机(SVM)、随机森林(RFC)、简单神经网络(NN)、集成学习($AdaBoost$)和极端梯度提升算法($XGBoost$)等,将周粒度下的电子台账数据转化为长向量引入模型;二维深度学习模型采用基于特征灰度的卷积神经网络($FG-CNN$),按测度对数据进行归一化并保持周粒度下电子台账数据的 15×52 矩阵形态,设置基于均方误差的早停策略防止学习模型过拟合。研究从预测准确性(预测精度)和稳定性(AUC)等两个方面评价上述预测方法。

3. 细分电子台账数据类别的预测能力对比。将电子台账数据划分为种植投入、电商交易和借贷保险,采用逐一剔除变量组的分组预测和绘制梯度加权类激活映射图两种方法揭示三类数据在风险预测上的作用差异。第一,分组预测构建不含种植投入(NPD)、不含电商交易(NSD)和不含借贷保险(NFD)三组不完整数据集进行风险预测,并将预测结果与基于完整数据集的风险预测结果进行对比;缺少某一类数据的预测结果下降最明显,则说明该类数据具有最强的预测能力,若同时使用三类数据的风险预测结果明显最优,则说明三类数据具有相互印证的信息含量。第二,梯度加权类激活映射图($Grad-CAM$)借助计算机视觉对深度学习结果的解释能力,将风险预测中更为重要的像素(即“时间—测度”坐标中的点)标记为更大的权重,从而直观反映出风险模式的形成位置。

三、变量统计与基础回归

(一) 样本采集与筛选

研究数据来自课题组对江苏省梨果种植户的调查,主要从以下4个方面筛选样本农户:第一,样本梨果种植户为家庭经营主体,且种植经营梨园面积不超过100亩;第二,样本农户以鲜果种植、加工和销售为主要经营范围,鲜果及加工产品收入占家庭总收入的50%以上(不限制梨果产品占比);第三,样本梨果种植户稳定参与农产品电商活动,电商收入占家庭收入10%以上;第四,剔除从事梨果种植和销售不足3年的农

户样本。最终采集3755名梨果种植户进入模型,以年为采样频率记录样本农户的经营特征,以周为采样频率记录样本农户的台账信息。研究以种植年(3月的第一个周一到来年3月的第一个周一的前一天)作为统计周期,2019年至2020年为农户经营数据和电子台账数据的统计年度,2020年至2021年为经营风险结果和决策数据的统计年度,从而验证对农产品经营风险的跨年短期预测。

(二) 变量的描述性统计

就农产品电商经营风险结果而言,在3755户样本梨果种植家庭中,32.01%(1202户)存在连续4个月及以上销售收入小于经营支出的现象,记为连续亏损(LK);22.31%(838户)存在发货或供货延期超过4周及以上的现象,记为流通受阻(LS);19.49%(732户)存在年还款额超过年销售总额(而非利润额)的现象,记为还款困难(HK)。就农产品电商经营风险决策而言,在3755户样本梨果种植家庭中,26.10%(980户)退出电商直销模式,记为直销退出(ZT);20.45%(768户)退出梨果家庭种植,记为种植退出(ST)。考虑到梨果种植连续性,转换种植品种的可能性较小,转出农地而从事专门的农业服务或脱离农业生产的可能性较大。

农户经营特征包含生产特征(农业劳动人数、土地流入比重、是否兼业经营)、电商特征(电商收入比重、电商投入摊销、电商订单比重)和环境特征(受灾减产比重、是否投保农险、是否经品牌认证)。生产特征结果表明样本家庭从事农业劳动的平均人数为2.85人,49.03%的梨园为流入土地,77.44%的样本家庭存在兼业经营。电商特征结果表明电商收入比重最高可达90.00%,平均43.75%的电商投入尚未摊销;同时样本中也无纯电商供货经营农户,参与电商活动的梨果种植户均存在电商直销行为和收入,为探讨电商风险下直销模式的退出提供了数据基础。环境特征结果表明农户平均受灾比重较低,但标准差达到23.74%,88.03%的样本农户参与农业保险;仅21.10%的农户能够在直销中使用获得认证的品牌。

电子台账数据包含种植投入(固定资产采购、易耗品采购、雇佣劳动支出、机具租赁支出、仓储管理支出)、电商交易(产品平均价格、产品销售数量、二次购买率、电商退货率、非销售收入)和借贷保险(赊购金额、赊销金额、信贷发生额、信贷还款额、保险赔付额)。种植投入结果表明样本梨果种植户在稳定生产下的年均固定资产采购费用仅219.21元,远低于年均易耗品采购的1805.23元。考虑部分易耗品以赊购形式获得,这一差距可能更大,同时雇佣劳动、机具租赁和仓储管理的年均费用支出均在千元以上。电商交易结果表明样本农户所售梨果的价格和销量差异较大,在不确定包装和物流成本时不能简单认为高售价和高销量能够带来高利润,新兴电商直播使部分农户获得广告收入和直播打赏等非销售收入,但在电商收入中仅占0.09%。借贷保险结果表明农户年均赊销和赊购差额超过6万元,信贷还款额低于信贷发生额,存在资金回笼困难。

(三) 基础回归结果

表1展示了农户经营数据对农业经营风险的影响结果,变量系数采用 *logistic* 模型估计;模型1至模型6以风险结果(连续亏损、流通受阻和还款困难)为因变量,模型7至模型10以风险决策(直销退出和种植退出)为因变量。模型中绝大多数变量系数均未通过10%水平下的显著性检验:第一,生产、电商和环境这三类特征片面解释了农产品经营风险。第二,将台账数据的年汇总值作为控制变量引入基础模型,则除了农险参保外,在不含控制变量的模型中具有显著影响的各项变量系数将不再显著,即引入台账数据挤出了农户经营数据估计系数的显著性,假说 H1a 成立。

(四) 线性预测结果

研究将两类数据分别引入 *logistic* 预测,表2结果表明假说 H1b 成立。以连续亏损、还款困难和种植退出为分类标记,仅使用台账数据的预测正确率明显高于仅使用特征数据的结果;其中,还款困难的预测结果显示,同时引入两类数据的预测正确率有可能低于仅使用台账数据结果。以流通受阻和直销退出为分类标记,仅使用统计数据的预测正确率略高于仅使用台账数据结果,使用两类数据、仅使用特征数据和仅使用台账数据的预测结果差异极小,说明两类数据在年汇总口径下具有高度的可替代性。但采用“台账数据 + *logistic*”的经营风险预测 *AUC* 均低于0.85,预测结果的稳定性和真实性较低,研究进一步引入非线性预测模型。

表1 基础回归结果

	模型1: LK	模型2: LK	模型3: LS	模型4: LS	模型5: HK	模型6: HK	模型7: ZT	模型8: ZT	模型9: ST	模型10: ST
农业劳动人数	-0.018 (-0.740)	0.021 (0.477)	-0.018 (-0.628)	0.004 (0.095)	0.002 (0.084)	0.069 (1.441)	-0.003 (-0.127)	0.031 (0.817)	0.005 (0.170)	-0.057 (-1.346)
土地流入比重	-0.580** (-2.355)	-0.381 (-0.877)	-0.144 (-0.516)	-0.302 (-0.727)	-0.530* (-1.843)	-0.461 (-0.950)	-0.551** (-2.155)	-0.315 (-0.804)	1.964*** (6.178)	0.634 (1.440)
是否兼业经营	0.097 (1.140)	-0.115 (-0.759)	0.195** (2.012)	0.164 (1.058)	0.189* (1.845)	0.184 (1.055)	0.171* (1.870)	0.224 (1.618)	-0.123 (-1.284)	-0.019 (-0.138)
电商收入比重	-0.044 (-0.351)	-0.260 (-1.227)	-0.078 (-0.551)	-0.281 (-1.319)	-0.126 (-0.841)	-0.397* (-1.694)	0.046 (0.346)	-0.004 (-0.020)	-0.048 (-0.327)	-0.010 (-0.045)
电商投入摊销	-0.239* (-1.805)	-0.099 (-0.456)	-0.174 (-1.176)	-0.006 (-0.028)	-0.192 (-1.227)	-0.017 (-0.070)	-0.210 (-1.492)	-0.069 (-0.354)	0.449*** (2.911)	0.278 (1.257)
电商订单比重	0.017 (0.092)	0.253 (0.809)	0.105 (0.497)	0.372 (1.201)	-0.128 (-0.583)	-0.016 (-0.046)	-0.017 (-0.086)	0.362 (1.274)	0.353* (1.653)	0.257 (0.813)
受灾减产比重	-0.101 (-0.674)	0.013 (0.038)	0.043 (0.259)	0.333 (1.052)	-0.055 (-0.317)	0.110 (0.312)	0.056 (0.354)	0.042 (0.153)	-0.085 (-0.474)	-0.037 (-0.113)
是否投保农险	0.095 (0.868)	0.101 (0.552)	0.047 (0.387)	-0.173 (-0.874)	0.016 (0.125)	-0.270 (-1.240)	-0.022 (-0.188)	-0.328* (-1.909)	0.134 (1.018)	0.190 (1.028)
是否经品牌认证	-0.045 (-0.502)	-0.172 (-1.090)	-0.022 (-0.219)	-0.062 (-0.405)	0.006 (0.055)	-0.169 (-1.013)	0.033 (0.354)	0.039 (0.289)	0.024 (0.234)	0.175 (1.199)
控制变量	不含	含	不含	含	不含	含	不含	含	不含	含
N	3755	3755	3755	3755	3755	3755	3755	3755	3755	3755

注:根据3755名梨果种植户调查数据整理;括号内为t值,*、**和***分别表示估计系数在10%、5%和1%水平下显著

表2 logistic 预测正确率(%)

解释变量	连续亏损	流通受阻	还款困难	直销退出	种植退出
两类数据	76.09	77.48	80.72	73.91	85.46
仅特征数据	68.00	77.69	80.51	73.91	79.55
仅台账数据	76.04	77.48	81.90	73.64	84.24

注:根据3755名梨果种植户调查数据整理

四、基于一维机器学习的经营风险预测

(一) 模型选择

研究引入 *SVM*、*RFC*、*NN*、*AdaBoost* 和 *XGBoost* 五种机器学习方法构建一维非线性预测模型。其中,支持向量机(*SVM*)是通过寻找最优分隔超平面对数据进行划分的算法,适合多变量、大样本下的非线性分类,由 Cortes 和 Vapnik (1995)^[31] 提出;随机森林(*RFC*)是在装袋法的基础上随机选择分裂变量的集成算法,可以将多个弱学习器集成成为强学习器,由 Breiman (2001)^[32] 提出;简单神经网络(*NN*)是引入少量隐藏层的感知机(Rosenblatt, 1958)^[33],经激活函数迭代后得到非线性分类结果,研究采用基于反向传播算法

(BP)的全连接神经网络(Rumelhart等,1986)^[34];自适应提升(AdaBoost)是通过修正观测值权重进行序贯分类调整的集成算法,最早由Freund和Schapire(1996)^[35]提出,后经Chen和Guestrin(2016)^[36]改进的极端梯度提升算法(XGBoost)更好地适应了大样本下的随机特征选择。

(二) 预测结果与分析

第一,如表3所示,连续亏损、流通受阻、还款困难和直销退出具有较强的非线性可预测特征,假说H2a成立。一方面,采用五种非线性算法的预测正确率在前四项风险指标上分别较基础预测增加11.18%、11.76%、8.41%和11.34%,说明农产品短期经营风险具有较强的非线性可分性。台账数据包含了农户对风险认知的有效认知并具备一种基于认知延续性的经营惯性,继而在跨年度的经营行为中表现出特定模式。这种模式不易被线性模型识别,但稳定地指向跨生产周期的经营风险,消除或缓解农户的经营风险应从优化农业经营的微观行为模式入手。另一方面,对种植退出的预测提升仅3.83%。梨果寿命较长,需4~5年坐果,而后进入40~50年的盛果期,因此梨果等林果产品具有较强的种植连续性和品种稳定性。将盛果期梨园承包出去并不是单纯的农地流转,这意味着过去一段时期的种植投入能够被合理补偿,而承包者往往也是看中种植的连片规模效应,而不会改变农地用途;因此研究也意识到电子台账可能遗漏决定农户退出梨果种植的一些突发性因素,如农业政策转变、平台资本注入和电商合作嵌入等外部激励因素。

表3 基于五种机器学习方法的预测正确率(%)

学习模型	连续亏损	流通受阻	还款困难	直销退出	种植退出
SVM(Radial Basis)	87.22	89.24	90.31	84.98	87.75
RFC	82.96	87.33	86.26	80.30	87.33
NN(Weight Delay)	83.71	86.48	88.39	77.00	86.69
AdaBoost	81.90	84.45	85.62	80.94	83.39
XGBoost	86.58	86.69	88.07	83.71	88.07
基于年汇总值的logistic	76.04	77.48	81.90	73.64	84.24

注:根据3755名梨果种植户调查数据整理

第二,非线性机器学习模型的预测表现存在异质性。SVM在前四项风险指标上的预测正确率较其他模型更高,将核函数由径向基函数调整为多项式或线性核则预测精度下降;说明通过向高维空间进行非线性映射可较为清晰地划分农产品经营风险的复杂分类边界。而对于种植退出等易受即时激励的复杂预测情境,SVM的表现稍逊于XGBoost,但调参耗时远远小于后者。考虑到直销和供货这两种电商模式并无优劣之分,对直销退出等风险决策的预测须将风险结果和农户电商认知习惯相结合。因此,预测风险决策的难度远远高于预测风险结果,这也是直销退出的预测精度明显低于三类风险结果的重要原因,后文将验证增加神经网络层数对预测结果的改善。

本节所采用的一维机器学习模型仍未充分利用台账信息在时间维度上的逻辑顺序,即表3所采用的方法都是将样本台账信息作为可任意旋转的向量引入模型,每个时间点上的变量信息是独立的,无法归纳出体现农户风险认知和行为习惯的序列模式。

五、基于二维深度学习的经营风险预测

(一) 模型选择

前述NN模型是包含2个隐藏层的全连接前馈神经网络,每个样本特征均表示为一个(780,)的长向量,因此第一个隐藏层的每个神经元都将与输入层关联,产生780个独立的神经连接;随着隐藏层及其神经元数量的增加,巨大的连接规模将显著降低神经网络的学习能力,极易造成过拟合的情形。因此,Hinton等(2012)^[37]认为在使用小数据集训练复杂前馈神经网络时,可通过随机省略特征检测器(Feature Detec-

tors)的方式避免过拟合,并在 AlexNet 的卷积神经网络中将 dropout 和全连接交替使用2次(Krizhevsky 等, 2017)^[38]。就农业生产的多维异质性而言,产品、品种和产地对农户生产经营行为为习惯的影响较大,训练数据集规模并不随采样范围的扩大而无限增长;当前3755个样本均为江苏省参与电商且具备台账记录基础的家庭梨果种植户,考虑林果种植的稳定性,采用 dropout 防止过拟合是构建这一规模下神经网络的必要选择。

如图2所示,农户的经营风险模式并非表现为单一类型或连续类型,其风险模式并不会出现在特定的采样时间或变量上,而是在两个维度上均相互抵消。因此,研究借鉴计算机视觉领域的二维卷积神经网络,将农户年度台账数据视为二维灰度图像,尝试捕捉风险模式的平移不变性和拉伸不变性。基于特征灰度的卷积神经网络(FG-CNN)采用有限像素的卷积核作为风险模式的过滤器,而多个卷积核所形成的卷积核可以抽取不同类型的风险模式;考虑经过卷积运算的特征维度仍较高,一般交替引入若干池化层以减少参数量、增加感知野,最终由全连接层训练输出。

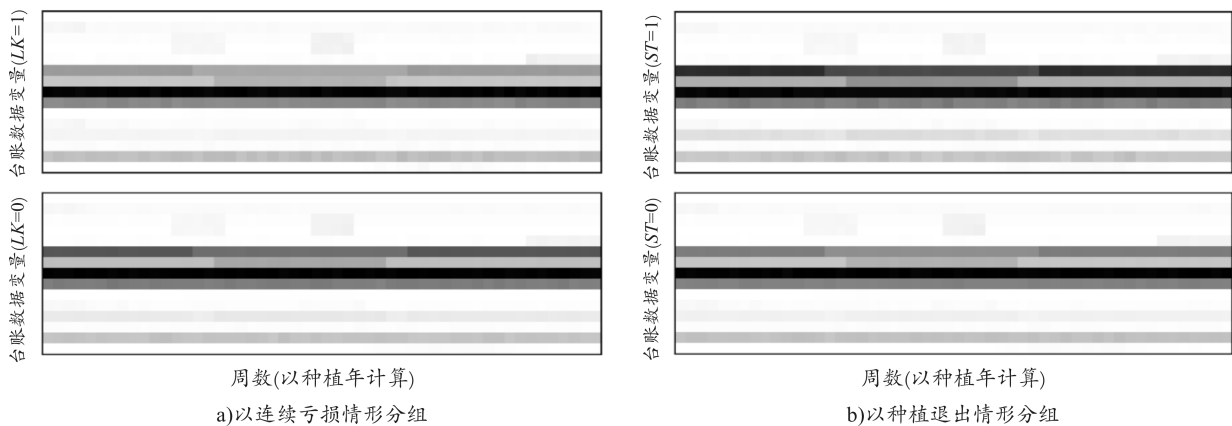


图2 分组展示的数据矩阵平均灰度图

说明:图中展示以连续亏损和种植退出为分类变量进行分组的不同分类下,15个台账变量与52周所形成的(15, 52)数据矩阵的平均特征灰度,其中图像灰度进行了翻转,颜色越浅表示归一化后的变量数值越大

(二) 预测结果与分析

研究采用(3,3)卷积核构建卷积网络并设置2个50%的 dropout 层,同时在验证集上建立基于均方误差的早停策略防止学习模型过拟合。图3和图4分别展示了风险结果和风险决策的深度学习预测结果,并绘制 ROC 曲线(黑色实线)与前述五种机器学习方法进行对比,基于特征灰度的卷积神经网络在预测准确性和稳定性上均超过一般机器学习方法,特别是对直销退出(ZT)指标的预测提升非常明显,假说 H2b 成立。

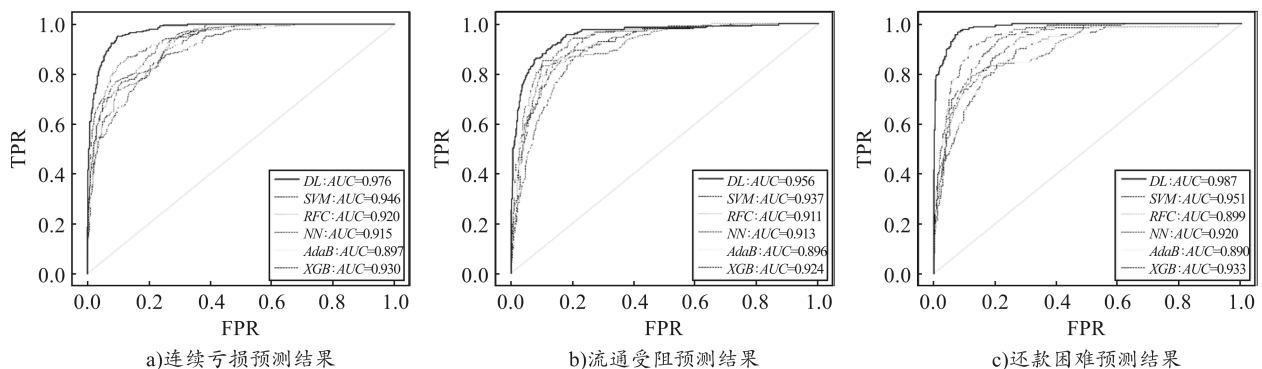


图3 风险结果预测的多方法对比

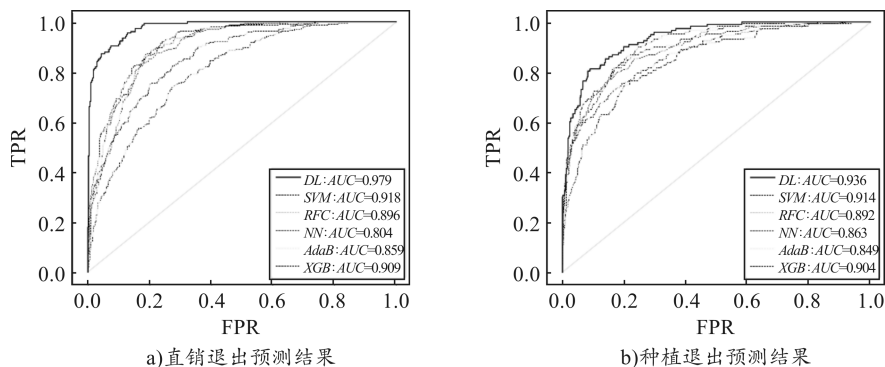


图4 风险决策预测的多方法对比

第一,除种植退出指标外,二维卷积神经网络明显提升了风险结果和决策的预测精确度。如表4第5列所示,风险决策中直销退出指标的预测精度提升非常明显,测试集预测正确率达到93.68%,较表3中SVM预测正确率84.98%提升了8.70%,并远远高于其他机器学习模型。将表4第2列至第4列与表3进行对比,三类风险结果指标的预测精度分别提升2.24%、1.39%和4.58%。一方面,表3中两个隐藏层的前馈神经网络预测精度未超过其他学习算法,但增加神经网络的隐藏层数量明显提升了训练效率,特别是在与农户经营风险认知和行为习惯相关的风险决策方面。说明二维深度学习能够从农产品经营行为中学习农户在风险决策上的一致性认知模式,从而增加对复杂结果的预测能力。另一方面,采用深度学习模型是将一维机器学习中长向量化的独立变量转化为具有时间逻辑的二维形态,并作为灰度图进行卷积处理,从而增加了有限感受野下的风险模式识别。

表4 卷积神经网络预测正确率(%)

数据集	连续亏损	流通受阻	还款困难	直销退出	种植退出
训练集(早停)	97.55	94.09	98.35	98.80	96.54
验证集(早停)	89.35	90.52	94.89	93.34	90.23
测试集	89.46	90.63	94.89	93.68	87.52

数据说明:根据3755名梨果种植户调查数据整理

第二,采用深度学习预测风险结果和决策时,样本外预测精度较高,小规模样本下的预测表现具有稳定性。对比表4中验证集和测试集的预测正确率可以看出,样本外预测(测试集)的正确率并不低于验证集,说明深度学习所识别的风险模式具有通用性,不存在明显的过拟合现象。从农户行为中识别出短期经营风险在时间维度上的通用模式,意味着风险预测可能存在跨期一致性,通过在时间维度上累积台账信息所形成的农户行为面板数据能够提升预测精度,从而对电商参与时间较短的同类农户建立更为有效的样本外预测模型。图3和图4的ROC曲线也表明,二维深度学习的预测结果较一维机器学习更为稳健,直销退出指标的深度学习AUC值为0.979,远高于SVM模型的0.918,说明二维预测在全部截断值下的预测精度要优于一维预测。

第三,虽然表4第6列种植退出的测试集预测正确率为87.52%,低于XGBoost算法的88.07%和SVM的87.75%,但其ROC曲线在一维机器学习算法之外,AUC值0.936也略高于XGBoost的0.901和SVM的0.914。其一,早停机制所给出的截断值并非全局最佳截断值,当相近轮次的正确率波动较大时,单一截断位置的正确率具有较高的随机性,其结果不如ROC曲线稳健;其二,种植退出的测试集正确率明显低于验证集,而在取消早停策略的图5中,样本外误差在轮次超过60后不降反升。说明台账信息具有较低的信噪比(Signal-Noise Ratio),农户的种植退出并不一定发生在两个种植年之间并易受外部激励影响,因此跨年预测相当于延长了其决策周期,忽略了即时信息对决策的短期影响,即时间维度上反向风险信号的抵消弱化了决策的冲动性,从而造成一定程度的过拟合(Rasekhschaffe和Jones,2019)^[39]。

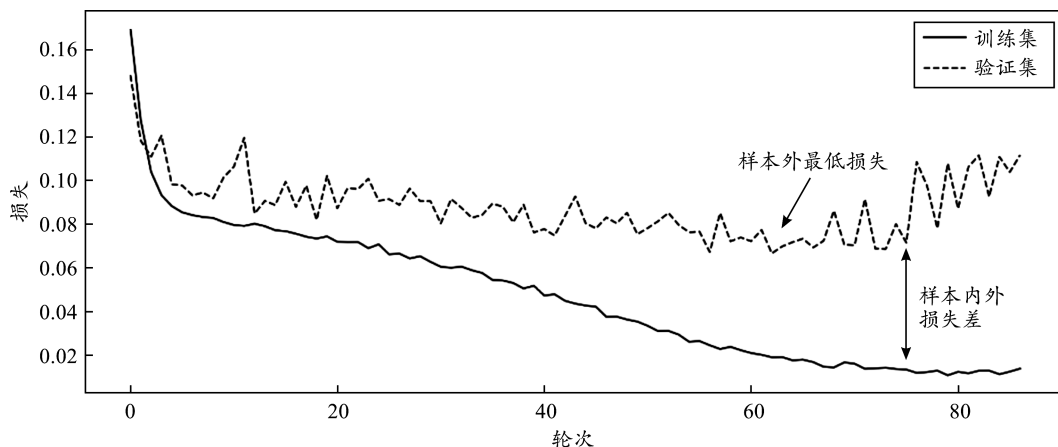


图5 取消早停策略的深度学习预测结果

(三) 三类行为信息预测能力差异分析

农户台账数据涉及种植投入、电商交易和借贷保险三个方面,考虑到三类变量的指标内涵、记录方式和信息精度存在差异,研究采用两种方法厘清不同行为信息对预测的贡献差异:一是逐一剔除变量组进行分组预测,预测精度下降越多则说明缺失的变量组对预测的贡献越大;这一方法缩小了信息规模并进一步降低了学习算法对判别信号的验证能力,从而高估了单一变量组对预测的贡献。二是借鉴计算机视觉的可解释性理论,采用梯度加权类激活映射图(Grad-CAM)定位算法感兴趣的判别区域(Zhou等,2015^[40]; Selvaraju等,2020^[41]);该方法通过热力图展示预测权重较大的二维区域,有效避免数据缺失对算法训练的影响,能够同时在时间和变量等两个维度上展示更细粒度的判别依据。

第一,如表5所示,三类行为信息均对农产品短期经营风险预测具有贡献,其中电商交易数据的贡献最大。剔除电商交易类变量后,连续亏损、流通受阻和还款困难的预测正确率分别下降15.66%、12.67%和14.59%,直销退出的预测正确率下降达19.77%,说明电商交易信息直接体现了农户的新渠道嵌入水平,形成农户短期经营预期。电商交易数据具有记录优势,逐笔交易的电子存根便于实时汇总和量化分析,其采样精度也远高于其他信息,可为后续研究构建更为复杂的预测模型提供良好的数据基础。同时,种植投入和借贷保险也具有不可忽视的贡献,剔除相应变量组均使风险预测精度下降。因此,三类台账数据在复杂预测模型中具有相互印证作用,剔除任意一类数据都将大幅降低预测精度,假说H3成立;直销退出指标下的分组预测ROC曲线如图6所示,其中全变量预测结果(DL)最优,剔除电商交易数据的预测结果(NSD)最差,剔除借贷保险数据的预测结果(NFD)略差于剔除种植投入数据的预测结果(NPD)。

表5 分组预测结果

模型划分	连续亏损	流通受阻	还款困难	直销退出	种植退出
不含种植投入数据	88.07	84.66	89.67	84.56	—
不含电商交易数据	73.80	77.96	80.30	73.91	—
不含借贷保险数据	86.26	87.01	91.27	84.66	—
含全部数据	89.46	90.63	94.89	93.68	87.52

注:根据3755名梨果种植户调查数据整理;种植退出指标在深度学习下的预测能力提升不明显,不再进行分类预测

第二,图7随机展示出两个直销退出样本的特征灰度图a)和变量激活强度热力图b)。激活强度最高的区域在电商交易变量组,并在时间维度具有平移不变性,说明电商交易所形成的数据对风险决策的预测贡献最大。有效权重的激活区域并不是一个特定的点,而是在整体上表现出以激活强度最高(颜色最深)的区域为中心向四周非线性衰减的形态。同时,借贷保险数据在种植年年末具有较高的预测贡献,激活中心位置对变量并不敏感;较为合理的解释是,农村金融的形式较为多样,农户习惯于赊购赊销和消费信贷等非现金收支的金融行为。但即便金融收支平衡,种植年年末也需要维护资金链以满足多主体间的销账和

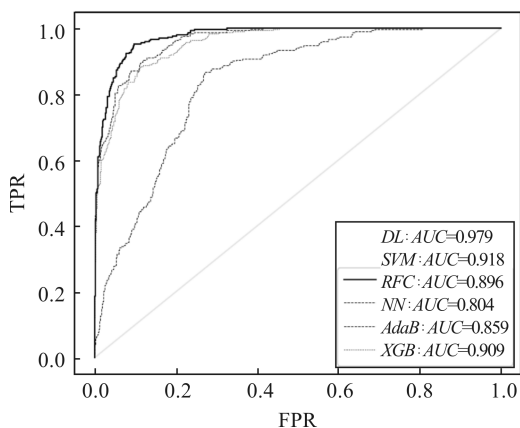


图6 完整数据与分组剔除数据的预测结果对比

还款,因此年末资金流动的价值明显且农户无须细分资金收支的具体项目,例如赊销资金回笼和农业保险赔付在风险预测视角下的作用是相同的。

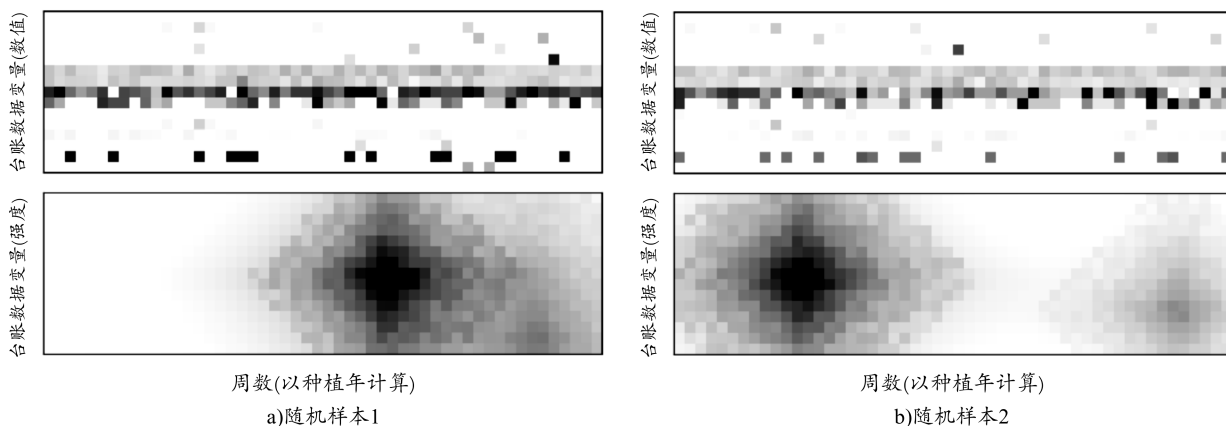


图7 梯度加权类激活映射图

六、结论与启示

通过对江苏省3755户梨果种植户的调查,获取农户经营数据和电子台账数据,对比分析线性模型、一维非线性机器学习模型和基于特征灰度的二维深度学习模型在预测农产品短期经营风险结果和风险决策上的差异。结果表明:第一,基于农户行为的电子台账数据较基于结果特征的农户经营数据更具准确预测农产品短期经营风险的信息含量,同时采用两类数据的风险预测结果并不明显优于仅采用电子台账数据的风险预测结果;第二,非线性机器学习方法较线性模型更能准确预测农产品电商经营风险,支持向量和 XGBoost 方法具有明显预测优势;第三,基于特征灰度的二维卷积神经网络提升了风险结果和决策的预测精准性和稳定性;第四,分组预测和梯度加权类激活映射图证实,电商交易信息具有最高的风险预测能力,借贷保险信息次之,种植投入信息的预测能力最差,且三者 in 风险预测中相互印证。

根据上述结论,研究得到以下四点启示:第一,农产品电子台账数据较传统农户经营特征数据更具预测风险的信息含量,因此规范农业经营数据的电子化形成和数字化记录是电商环境下提升农户经营能力、优化农户参与模式的重要手段,应开发和完善农业经营台账管理系统,做到与农产品电商过程高度适配,鼓励各类电商服务主体嵌入数据平台,形成可相互验证的多维电商大数据。第二,机器学习方法在农产品短期经营风险预测的准确性和稳定性上超过传统线性预测模型,更适合处理高粒度下的电商大数据,且支

持向量机和 *XGBoost* 等方法在不同风险结果和风险决策的预测上各具优势,因此在开发农业电子台账平台时,可引入多种基于机器学习的农产品经营风险预测算法,对农户当前的经营状况进行实时预警,帮助农户及时调整经营策略。第三,基于特征灰度的二维卷积神经网络在一维机器学习的预测结果基础上又有大幅提升,但对算力的要求较高,适用于电商监管部门和有条件的电商服务部门对参与电商经营的农户进行风险识别和精准画像,同时应引导建立和完善第三方农业电商经营咨询机构,吸引具备数字技能的农户向农业服务业转移,从事专门的农业数字化管理和规划服务,构建科学的农业资源托管模式。第四,种植投入、电商交易和借贷保险数据具有相互印证的风险预测能力,农业服务部门应构建自动化的电商大数据多维验证机制,进一步形成和完善微观农业数据要素利用场景,助力小农户以软信息嵌入电商供应链,从而突破固有资源禀赋限制。特别是当前种植投入数据的采集手段粗糙,往往需要农户主动人为记录,造成数据误差大、预测能力弱,因此迫切需要加快农村物联网和云服务建设,为农户记录和管理采购、运输、仓储等服务数据提供便利。

受限于数据质量和采样规模,研究仍存在两个方面的不足:第一,为兼顾不同采样渠道下的数据粒度而选取以周为统计频率并不能充分体现电商数据的高频性。日汇总数据、分时汇总数据和逐笔交易数据可能具有更高的预测信息含量,但更高的采样频率对预测模型的选取和调参提出了更高的要求。当频率足够高且同步增量处理时,当前模型通过固定时间维度而汇总行为信息的数据透视结构将不再适用,仍须验证农产品短期经营风险是否由特定环境下的经营行为触发。第二,当前由电子台账数据构建的二维数据结构仅包含单一种植周期,无法精准识别农户的长期经营惯性和风险偏好,弱化了农户经营脆弱性认知和电商组织化合作的异质性,因此在种植退出等复杂风险决策上未表现出明显的预测优势。而提升采样频率、叠加采样周期将综合识别种植年内行为序列和种植年间行为差异,后续研究仍须进一步验证其信噪比变化对预测精确性和稳定性的影响。

参考文献:

- [1] 陈新建,韦圆圆. 风险感知、风险偏好与贫困农户风险管理策略[J]. 华南农业大学学报(社会科学版),2019(1):74-85.
- [2] 沈琼. 新型职业农民持续务农的关键因素及面临的难题[J]. 社会科学家,2020(9):15-21.
- [3] 尚燕,熊涛. 所为非所想? 农户风险管理意愿与行为的悖离分析[J]. 华中农业大学学报(社会科学版),2020(5):19-28,169.
- [4] 丁少群,张珏,李丹. 乡村振兴背景下农业巨灾风险分散机制完善研究[J]. 理论探索,2021(5):96-104.
- [5] 王志刚,李圣军,宋敏. 农业收入风险对农户生产经营的影响:来自西南地区的实证分析[J]. 农业技术经济,2005(4):46-50.
- [6] 谢花林,黄莹乾. 非农就业与土地流转对农户耕地撂荒行为的影响——以闽赣湘山区为例[J]. 自然资源学报,2022(2):408-423.
- [7] PORTER M. Strategy and the internet[J]. Harvard Business Review,2001,79(3):63-78.
- [8] KHANAL R, MISHRA K, GUJARATHI P, et al. Financial performance of small farm business households: the role of internet [J]. China Agricultural Economic Review,2016,8(4):553-571.
- [9] 周浪. 另一种“资本下乡”:电商资本嵌入乡村社会的过程与机制[J]. 中国农村经济,2020(12):35-55.
- [10] DUTTA P, SURYAWANSHI P, GUJARATHI P, et al. Managing risk for e-commerce supply chains: an empirical study [J]. IFAC-PapersOnLine,2019,52(13):349-354.
- [11] ZHANG N, WEI X. Game theory analysis on credit risk assessment in e-commerce [J]. Information Processing & Management, 2022,59(1):1-9.
- [12] 肖迪,陈瑛,王佳燕,等. 考虑平台数据赋能的电商供应链成本分担策略选择研究[J]. 中国管理科学,2021(10):58-69.
- [13] 汪兴东,刘雨虹. 农户电商销售意愿与行为偏差分析——基于江西省54个县的调查数据[J]. 农林经济管理学报,2021(3):316-325.
- [14] 王艳玲,张广胜. 农户电商创业意愿实证研究——基于技术接受、感知风险与主观经验[J]. 北京交通大学学报(社会

- 科学版),2021(2):90-100.
- [15] 向丽,胡珑瑛. 生计风险感知对农民参与电商扶贫意愿的影响及代际差异研究[J]. 农业技术经济,2019(5):85-98.
- [16] 郝鸿,刘尊礼,孟突忠,等. 行使权力对渠道电商感知风险的调节作用——感知关系风险和绩效风险的差异[J]. 工业工程与管理,2015(5):37-44,53.
- [17] 聂召英,王伊欢. 链接与断裂:小农户与互联网市场衔接机制研究——以农村电商的生产经营实践为例[J]. 农业经济问题,2021(1):132-143.
- [18] 廖理,崔向博,孙琼. 另类数据的信息含量研究——来自电商销售的证据[J]. 管理世界,2021(9):90-104.
- [19] KAMBLE S, GUNASEKARAN A, GAWANKAR S. Achieving sustainable performance in a data-driven agriculture supply chain [J]. *International Journal of Production Economics*, 2019(5):179-194.
- [20] 林明,杨林楠,彭琳,等. 基于BFGS-NARX神经网络的农产品价格预测方法[J]. 统计与决策,2013(16):18-20.
- [21] 王骞. 基于遗传算法与支持向量机的农业经济预测[J]. 农业技术经济,2011(8):34-38.
- [22] 谢星雨,余本功. 基于MFMB的电商评论文本分类研究[J]. 数据分析与知识发现,2022(1):101-112.
- [23] 岑咏华,李文敬,李莉,等. 基于深度学习的电商评论信息多刻面情感分类研究[J]. 情报科学,2021(9):67-73,100.
- [24] 胡春华,邓奥,童小芹,等. 社交电商中融合信任和声誉的图神经网络推荐研究[J]. 中国管理科学,2021(10):202-212.
- [25] 王颖,阮梦黎. 基于大数据的电商异常交易风险评估仿真[J]. 计算机仿真,2018(3):369-372,388.
- [26] 徐鲲,张驰,鲍新中. 基于电商平台自有资金的供应链融资风险系统分析与仿真[J]. 系统科学学报,2021(2):100-104,115.
- [27] DAI N, VIKEN G, JOO E, et al. Risk assessment in e-commerce: how sellers' photos, reputation scores, and the stake of a transaction influence buyers' purchase behavior and information processing[J]. *Computers in Human Behavior*, 2018, 84(7):342-351.
- [28] 刘艳. 中国农业弱质性探究[J]. 求是学刊,1998(3):47-50.
- [29] ROUMASSET J. Rice and risk: decision making among low-income farmers[M]. Amsterdam: North-Holland, 1976:825-828.
- [30] 程欣炜,岳中刚. 参与农产品电商能够提升农村女性有薪劳动绩效吗?——基于10省种植户家庭跟踪数据[J]. 商业经济与管理,2021(10):5-19.
- [31] CORTES C, VAPNIK V. Support vector networks[J]. *Machine Learning*, 1995, 20(3):273-297.
- [32] BREIMAN L. Random forests[J]. *Machine Learning*, 2001, 45(1):5-23.
- [33] ROSENBLATT F. The perceptron: a probabilistic model for information storage and organization in the brain[J]. *Psychological Review*, 1958, 65(6):386-408.
- [34] RUMELHART D, HINTON G, WILLIAMS R, et al. Learning representations by back propagating errors[J]. *Nature*, 1986, 323(6):533-536.
- [35] FREUND Y, SCHAPIRE R. Experiment with a new boosting algorithm[A]//*Machine Learning: Proceedings of the Thirteenth International Conference*. San Francisco: Morgan Kaufman, 1996:148-156.
- [36] CHEN T, GUESTRIN C. XGboost: a scalable tree boosting system[C]. San Francisco: 22nd SIGKDD Conference on Knowledge Discovery and Data Mining, 2016.
- [37] HINTON G, SRIVASTAVA N, KRIZHEVSKY A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. *Computer Science*, 2012(4):212-223.
- [38] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6):84-90.
- [39] RASEKHSCHAFFE K, JONES R. Machine learning for stock selection[J]. *Financial Analysts Journal*, 2019, 75(3):70-88.
- [40] ZHOU B, KHOSLA A, LAPEDRIZA A, et al. Learning deep features for discriminative localization[J]. ArXiv Preprint ArXiv:1512.04150, 2015:1-10.
- [41] SELVARAJU R, COGSWELL M, VEDANTAM R, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[J]. *International Journal of Computer Vision*, 2020, 128(2):336-359.

